P22, L, and herpesvirus may have similar folds, suggesting a common capsid protein ancestor (Baker et al., 2005). A common ancestry is also suggested for phage PRD1 and adenoviruses based on structural studies (Bamford et al., 2005). Continued investigation of phage and their eukaryotic homologs is likely to reveal more common structural and functional ancestry amongst these divergent viruses.

Carolyn M. Teschke[1]
[1] Department of Molecular and Cell Biology
University of Connecticut
Storrs, Connecticut 06269

Selected Reading

Baker, M.L., Jiang, W., Rixon, F.J., and Chiu, W. (2005). J. Virol. *79*, 14967–14970.

Bamford, D.H., Grimes, J.M., and Stuart, D.I. (2005). Curr. Opin. Struct. Biol. *15*, 655–663.

Caspar, D.L.D., and Klug, A. (1962). Cold Spring Harb. Symp. Quant. Biol. *27*, 1–24.

Gilcrease, E.B., Winn-Stapley, D.A., Hewitt, F.C., Joss, L., and Casjens, S.R. (2005). J. Bacteriol. *187*, 2050–2057.

Heymann, J.B., Cheng, N., Newcomb, W.W., Trus, B.L., Brown, J.C., and Steven, A.C. (2003). Nat. Struct. Biol. *10*, 334–341.

Parent, K.N., Zlotnick, A., and Teschke, C.M. (2006). J. Mol. Biol., in press.

Parks, R.J. (2005). Mol. Ther. *11*, 19–25.

Sternberg, N., and Weisberg, R. (1977). J. Mol. Biol. *117*, 733–759.

Steven, A.C., Heymann, J.B., Cheng, N., Trus, B.L., and Conway, J.F. (2005). Curr. Opin. Struct. Biol. *15*, 227–236.

Tang, L., Gilcrease, E.B., Casjens, S.R., and Johnson, J.E. (2006). Structure *14*, this issue, 837–845.

Teschke, C.M., McGough, A., and Thuman-Commike, P.A. (2003). Biophys. J. *84*, 2585–2592.

Yang, F., Forrer, P., Dauter, Z., Conway, J.F., Cheng, N., Cerritelli, M.E., Steven, A.C., Pluckthun, A., and Wlodawer, A. (2000). Nat. Struct. Biol. *230*, 230–237.

Structure *14*, May 2006 DOI 10.1016/j.str.2006.05.001

Broken Symmetry in Homing Endonucleases

Homing DNA endonucleases are highly site-specific enzymes that initiate the transfer of mobile DNA elements. In this issue of *Structure*, Spiegel et al. report the structure of the I-CeuI homing enzyme and describe how a symmetric homodimeric enzyme acquired specificity for an asymmetric substrate.

Paleontologists gradually piece together the evolution of a dinosaur species, revising their model as new fossils are discovered. If only structural biologists were as fortunate. There is no readily available "fossil record" of extinct progenitor proteins to shed light on the evolution of the proteins that are observed today. Instead, structural biologists scrutinize the structures of modern proteins from different species to discern a plausible evolutionary scenario. The newly described structure of the I-CeuI homing endonuclease by Stoddard and coworkers (Spiegel et al., 2006 [this issue of *Structure*]) is used in this approach to help explain the extensive proliferation of homing endonuclease genes (HEGs) in nature.

Homing endonuclease genes are mobile selfish DNAs that have been remarkably successful invaders of diverse genomes. HEGs propagate by homing, a biased gene conversion event that duplicates the HEG to a homologous recipient allele that lacks the element. The HEG encodes a site-specific endonuclease that cleaves a target site located within the recipient allele, thereby stimulating DNA repair. Repair that uses the HEG-containing allele as a template leads to transfer of the genetic element (Stoddard, 2005). HEGs are usually found within introns or inteins, rendering them invisible to the host as they are excised at the RNA or protein level. An evolutionary cycle has been proposed for HEGs involving horizontal transmission to a naive genome, spreading and fixation in the recipient population by homing, degeneration of the HEG once all recipient alleles are converted to donors, and eventual loss from the genome (Goddard and Burt, 1999). The cycle is repeated if the element reenters the same site, moves to a related site in the genome, or horizontally transfers to a related site in the genome of a different host species.

How do homing endonucleases hone their specificity during evolution such that they cleave a single target within a complex genome to initiate homing while maintaining the site-recognition flexibility that permits the HEG to invade new genomes? Examination of several structures of LAGLIDADG homing enzymes, which comprise the largest family, provides some of the answers. LAGLIDADG enzymes utilize an extended protein-DNA interface covering up to 31 base pairs to acquire their necessary specificity. However, within this interface, homing endonucleases make contact to only a limited subset of the total available hydrogen bond partners in the DNA. For example, I-SceI forms hydrogen bonds to only 40% of the potential contacts in its target site (Moure et al., 2003). Consequently, homing endonucleases tolerate extensive base variation within their recognition sequence, which facilitates their movement to target sites in the same or other genomes that have undergone genetic drift.

Structural features of LAGLIDADG enzymes suggest a hypothetical evolutionary pathway. The simplest enzymes observed today are homodimers, such as I-CreI, that recognize palindromic or pseudopalindromic target

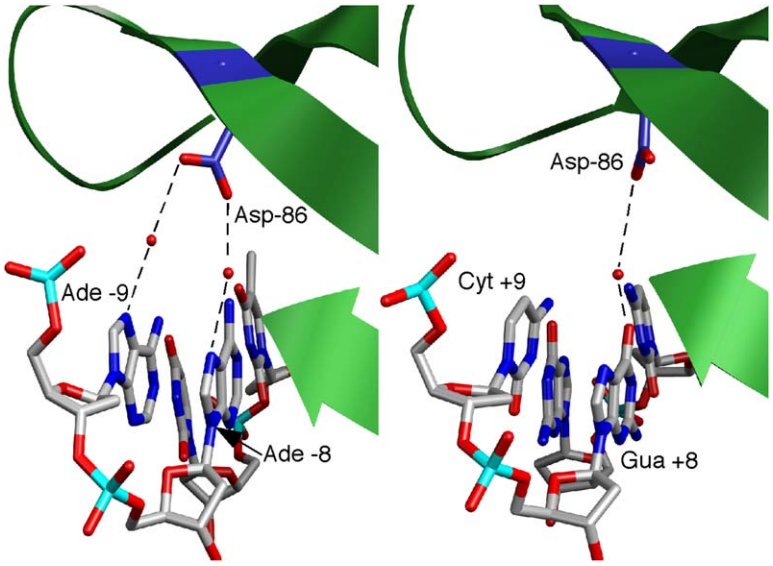

Figure 1. Asp-86 Adopts Different Conformations in Each I-CeuI Subunit

Asp-86 in one I-CeuI subunit (left) makes water-mediated contacts to two A:T base pairs at positions −9 and −8 of the I-CeuI recognition sequence. In contrast, the same residue in the second subunit of the homodimer (right) makes a single water-mediated contact to a G:C base pair at position +8, but does not contact the C:G base pair at +9. These asymmetric contacts result because the two aspartic acids utilize different rotameric torsion angles. Water molecules are shown as red spheres. The coordinates used to depict the structures in the figure (adapted from Spiegel et al. [2006]) were provided by Clint Spiegel and Barry Stoddard.

sites. These likely evolved from single subunit monomeric proteins that have no present-day counterparts. For homing endonucleases, as for most type II restriction enzymes, the advantage to forming homodimers was that they acquired increased DNA binding affinity through cooperative effects and the ability to recognize a DNA target sequence that is doubled in length (Marianayagam et al., 2004). Gene duplication is thought to have generated monomeric homing enzymes such as I-SceI that contain two LAGLIDADG motifs and recognize completely asymmetric targets (Lykke-Andersen et al., 1996). The ability of homing enzymes to contact asymmetric sequences increases their flexibility in evolving target-site specificity. Subsequent elaboration of monomeric enzymes with small polypeptide modules and invasion of a protein splicing gene is thought to have led to intein-encoded homing enzymes like PI-SceI (Duan et al., 1997).

How homing enzymes evolved from recognizing symmetric to asymmetric targets is unclear, but the new I-CeuI structure provides some clues (Spiegel et al., 2006). I-CeuI, like I-CreI and the related enzyme I-MsoI (38% identity), is a homodimer, and all three enzymes have similar overall topologies, but I-CeuI is only 15% identical to those enzymes and is part of a different subfamily, since the group I intron in which its ORF is located is inserted at a different site within the host gene. What also sets I-CeuI apart is the higher degree of asymmetry of its ~22 base pair target site, since it is significantly less palindromic (36%, 4/11 base pairs per half-site) than that of I-CreI (64%) or I-MsoI (45%). Interestingly, I-CeuI cleaves its asymmetric target faster than either of two synthetic palindromic targets that are comprised of two copies of either half-site. The asymmetric sequence may accommodate a bent DNA conformation required for protein binding or for catalysis better than either palindromic site (Spiegel et al., 2006).

The structural basis for the recognition of the asymmetric target was visualized in the high-resolution crystal structure (Spiegel et al., 2006). In each I-CeuI subunit, rotationally symmetric side chains assume different rotameric torsion angles and use bridging water molecules differently in order to contact nonidentical base pairs at the same position of each half-site. For example, the two carboxylate oxygens of Asp-86 in one I-CeuI subunit use water-mediated hydrogen bonds to contact the adenine N3 nitrogen atoms in consecutive A:T base pairs located at positions −8 and −9 of the left half-site (Figure 1). In contrast, in the right half-site, the symmetry-related aspartate side chain is rotated 90° about its Chi(2) torsion angle and makes a single water-mediated contact to the extracyclic oxygen of guanine of a G:C base pair at position +8 and makes no contact to a C:G base-pair at position +9. The significance of this finding is that it suggests that there was sufficient flexibility inherent in each protein subunit to recognize a predominantly asymmetric target even before gene duplication occurred to form monomeric enzymes. Thus, not only do homing endonucleases facilitate HEG transfer by selecting different bases to contact, but also by using alternative residue conformations to contact the bases.

Although the homing endonuclease progenitors may be gone forever, some of the putative intermediates in the evolutionary pathway have been reconstructed through protein engineering. The tethering of two I-CreI subunits to generate a single-chain monomer recapitulates one way that homodimer enzymes can give rise to dual domain monomers (Epinat et al., 2003). Yet to be found is a heterodimeric form of an enzyme in which each subunit diverged in specificity following gene duplication. The heterodimeric restriction enzymes that recognize asymmetric target sites may be analogous to this putative enzyme intermediate (Bellamy et al., 2005). Without waiting for such a homing endonuclease to be found in nature, two groups engineered an artificial heterodimer by combining subunits from I-DmoI and I-CreI, yielding an enzyme with novel specificity that has potential applications in genomic engineering (Chevalier et al., 2002; Epinat et al., 2003). As more structures of homing endonucleases become available, more clues to the remarkable evolutionary history of homing

elements may be uncovered, and more advances in enzyme design will be possible.

Frederick S. Gimble[1]
[1] Department of Biochemistry
175 South University Street
Purdue University
West Lafayette, Indiana 47907

Selected Reading

Bellamy, S.R., Milsom, S.E., Scott, D.J., Daniels, L.E., Wilson, G.G., and Halford, S.E. (2005). J. Mol. Biol. *348*, 641–653.

Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J., and Stoddard, B.L. (2002). Mol. Cell *10*, 895–905.

Duan, X., Gimble, F.S., and Quiocho, F.A. (1997). Cell *89*, 555–564.

Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F., and Lacroix, E. (2003). Nucleic Acids Res. *31*, 2952–2962.

Goddard, M.R., and Burt, A. (1999). Proc. Natl. Acad. Sci. USA *96*, 13880–13885.

Lykke-Andersen, J., Garrett, R.A., and Kjems, J. (1996). Nucleic Acids Res. *24*, 3982–3989.

Marianayagam, N.J., Sunde, M., and Matthews, J.M. (2004). Trends Biochem. Sci. *29*, 618–625.

Moure, C.M., Gimble, F.S., and Quiocho, F.A. (2003). J. Mol. Biol. *334*, 685–695.

Spiegel, P.C., Chevalier, B., Sussman, D., Turmel, M., Lemieux, C., and Stoddard, B.L. (2006). Structure *14*, this issue, 869–880.

Stoddard, B.L. (2005). Q. Rev. Biophys. *38*, 49–95.

Structure *14*, May 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2006.05.002

Turning up the HEAT on Translation

Eukaryotic Initiation Factor 4G forms the core of the translation initiation complex. Bellsolell et al. report the structure of its C-terminal region as two HEAT domains that define a new class and map binding sites for its regulatory factors (Bellsolell et al., 2006).

Initiation of mRNA translation in eukaryotes is subject to complex regulation and involves a large number of initiation factors (eIFs) that coordinate recruitment of 5′-capped mRNA and the ribosomal subunits (Figure 1A). This control is valuable for eukaryotes as it can be localized in cells and act very rapidly, without the need to wait for regulation at the transcriptional level. In some cases, transcription may even be switched off completely, such as in reticulocytes (immature red blood cells), where globin production must be regulated but the nucleus no longer functions, and in early development. Errors in translational regulation, such as those caused by mutation in eIFs, are found in various human diseases, especially cancers (Abbott and Proud, 2004; Watkins and Norbury, 2002).

Cellular eukaryotic mRNAs are capped at their 5′ terminus with a modified base (m^7G) and 5′-5′ phosphate linkage, with the exception of those in organelles. eIF4E binds the cap and carries it to the core scaffolding protein eIF4G (Figure 1A). The poly(A) tail of the mRNA is bound by poly(A) binding protein (PABP), which is itself bound by eIF4G, circularizing the message. eIF4G also binds MAP kinase signal-integrating kinase 1 (Mnk1) that regulates eIF4E activity by phosphorylation. An ATP-dependent helicase, eIF4A, also binds to eIF4G and serves to unwind double-stranded RNA to help expose the AUG start codon. The complex of eIF4A, eIF4G, and eIF4E is known as eIF4F (Gingras et al., 1999). eIF2, in the GTP bound form, recruits the initiator methionyl-tRNA (Met-$tRNA_i^{Met}$) to the 40S subunit of the ribosome to form a ternary complex, which is in turn recruited to eIF4F by eIF3, which binds both the 40S subunit and eIF4G. The Met-$tRNA_i^{Met}$ then scans the message to find the initiator AUG codon, at which point translation can begin.

In humans, eIF4G is a large protein of over 1500 amino acids that can be cleaved into three approximately equal regions by picornavirus proteases. The N-terminal region binds eIF4E and PABP, while the middle region (4G/M) has binding sites for eIF4A and eIF3. These regions are essential for cap-dependent translation and separating them by cleavage is inhibitory. Viral proteases carry out this cleavage in order to favor translation of the viral mRNA, which is uncapped and initiates from an internal ribosomal entry site. The C-terminal region (4G/C) is only present in higher eukaryotes, but not in yeast or *C. elegans*, and has additional regulatory functions, with binding sites for Mnk1 and eIF4A. For instance, recruitment of Mnk1 stimulates cap-dependent translation as a result of phosphorylation of eIF4E. In this issue, Bellsolell et al. report the structure of the C-terminal region of the scaffolding protein eIF4G from humans.

Previous work by the Burley group showed that the structure of the 4G/M region contains a typical α-helical HEAT-repeat (HR) domain (Marcotrigiano et al., 2001). HEAT repeats were first identified by Andrade and Bork (1995) in *H*untingtin, *E*F3, PR/65*A* subunit of PP2A, and *T*or1 and are common in proteins associated with translation. Each HR consists of two helices joined by a short loop, and they typically pack sequentially with interrepeat twist angles of +20°, with occasional values of about −45°, typically resulting in extended superhelical structures. Figure 1B shows the structure of PP2A (Groves et al., 1999), which has 15 HRs, with the first five in a similar arrangement to the HRs in 4G/M. In contrast, 4G/C contains two HEAT domains, 4G/C1 and 4G/C2, where *all* the interrepeat twist angles are about −45° (compare Figure 4A to Figure 4B in Bellsolell et al.). This unexpected arrangement results in a novel fold, which is much more compact than other known HEAT domains and represents a new subclass. Burley and coworkers go on to identify specific clusters of hydrophobic residues that favor this arrangement and